KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

A decision support system for the stochastic aggregate planning problem

Lieckens K, Vandaele N.

KBI_1406

A decision support system for the stochastic aggregate planning problem

Kris Lieckens*[1] and Nico Vandaele†[1,2]

[1]Faculty of Business and Economics, Research Center for Operations Management, KU Leuven, Belgium

[2]Faculty of Business and Economics, Kulak, Belgium

March 28, 2014

Abstract

An advanced decision support system is presented to answer aggregate planning questions regarding the trade-off between demand (product-mix) and supply (capacity) in a multi period stochastic setting. This tool improves the effectiveness and efficiency of sales and operation planning meetings by accounting for both revenues and costs that are relevant at the intermediate planning horizon. We develop a multi product, multi routing model, where a routing consists of a sequence of operations on different resources. Given customer demand in each time period, the model obtains the optimal production quantities in every period for each alternative routing, while explicitly taking into account the stochastic nature of both demand patterns and production lead times. This is the key difference between our approach and traditional aggregate planning models. At the same time, an optimal capacity level for each resource is obtained. We include trade-offs between level and chase strategies by

*kris.lieckens@kuleuven.be (corresponding author)
†nico.vandaele@kuleuven.be

charging costs for inventory, work-in-process, backorders, setups, regular time, overtime, etc. Outsourcing is considered as an alternative source with a stochastic lead time. The methodology builds upon a queueing network to estimate product's lead time distribution and associated quoted lead time with a service level. More system improvements can be obtained by proper lot sizing. This model is a mixed integer non-linear programming problem. We show that the search process of the differential evolution algorithm is efficient to find stable results within acceptable time limits. A scenario analysis reveals interesting managerial insights.

Keywords: Aggregate production planning, Sales and operations planning, Queueing network, Differential evolution

1 Introduction

Each member in a supply chain system, regardless of its focus (production, service, transportation or warehousing), has to perform two functions with respect to planning the respective business processes, namely (1) finding the balance between supply and demand and (2) link different strategic and operational plans through cross-functional integration. Both functions are the main purpose of the sales and operations planning (S&OP). According to the APICS dictionary, this process develops tactical plans that provide management the ability to strategically direct its businesses to achieve competitive advantage on a continuous basis by integrating customer-focused marketing plans for new and existing products with the management of the supply chain (Blackstone, 2010). It is typically performed on a monthly basis and is reviewed by management at an aggregate, product family level. It is the definitive statement of the company's plans for the near to intermediate term covering a horizon sufficient to plan for resources and support the annual business planning process. Since this process integrates all the plans for the business (sales, marketing, development, manufacturing, sourcing, and financial), it is important to outline the focus of this research. We want to build a manufacturing plan that satisfies several criteria:

– it is expressed in financial terms

– it has a mid term planning horizon (typically 12 or 18 monthly time buckets)

– it is based on both confirmed sales and forecasted demand in each time bucket for multiple products and various family types

– it optimally assigns the product-mix to the resource capacities in each time bucket

– it determines the optimal mix of in-house and outsourced production (make or buy decision)

– it determines the best allocation of in-house production to the alternative routings

– it determines lead time offsets with a given service level (lead time quotes) for each product in each time bucket

– it determines the optimal lot size for each product in each time bucket

These are typical managerial issues addressed by aggregate production planning (APP) models, which are situated at the tactical, mid-term rough-cut-capacity planning level and whose output determines the master production schedule. However, standard APP models usually neglect the flow dynamics and the stochastic behavior of the underlying system, the load dependent lead time effects and the lot size impact on the lead time. However, all these factors are prominent present not only due to fluctuations in demand and capacity levels, but also due to mix changes and stochastic elements in processing. This makes it difficult for standard APP models to produce robust production plans that can be fulfilled with a high level of customer satisfaction. To this end, we include steady state equations from queueing theory into the objective function in order to account for congestion and variability effects at the aggregate planning level. At this level, the time buckets in the planning horizon are long enough to ensure steady state conditions. The obtained lead time distribution enables us to estimate the service level that corresponds to the proposed lead time offset, i.e. time delay used to determine the release time when the completion time is given by the market, or equivalently, the delivery time to promise to customers. A fundamental queueing relationship states that lead time increases nonlinearly in both mean and variance with resource utilization, which is system's workload relative to its capacity (Hopp and Spearman, 2000).

More specifically, the model is developed for supply chain and production systems where multiple products, possibly grouped into family types, are obtained either from an external subcontractor or from internal production. In case of the latter, the appropriate capacity level must be set for various resources, i.e. machines or labor with one or multiple parallel servers that are used to perform a product dependent sequence of operations. Different sequences of process steps, also referred to as alternative routings, may exist for each product, while scrap and rework may arise at each operation. Furthermore, demand volume in each time bucket must be assigned to one of the time buckets in the planning horizon where the actual production will take place. If this demand quantity is split and some volume is produced earlier than required, inventory arises. If it is produced later than required, backorders are created. The utilization level, which is the result of confronting lot size dependent workload with available capacity, is the main determinant of the congestion effects, and thus of a product's total lead time. Nevertheless, a significant contribution to these delays can also be expected from variabilities and disruptions caused by uncertain demand intervals, unstable setup and process times, quality problems (scrap, rework, breakdowns, ...), lot sizing, various working schedules, etc. Since the model that we propose in this paper captures all these effects, it is denoted as the Advanced Resources Planning (ARP) as introduced in Vandaele and De Boeck (2003) and Nieuwenhuyse et al. (2011).

With given costs for work-in-process (WIP), finished goods, backorders, setups, production, disposal, regular time, overtime and outsourcing on the one hand and a fixed selling price for each product on the other, the goal is to maximize profit while taking into account customer service level with respect to due date delivery performance. Since Thomé et al. (2012) have found that lead times, capacity utilization, costs, quality, flexibility, variability, shortages and backlogs, timeliness and reliability of deliveries, and user satisfaction are all key metrics in the S&OP, the ARP model will produce production plans that are more realistic and more cost efficient when compared to the outcome of standard APP models. In contrast to these APP models where lead times, lot sizes and capacity buffers are fixed and predetermined, the ARP model dynamically adapts these parameters over time in function of S&OP decisions concerning production volume, capacity levels and the leanness of process characteristics. This accommodates for a more gradual recovery from losses due to e.g. setups, resource breakdowns and other production delays. It is commonly known that models not incorporating these stochastic elements are prone to inferior planning decisions when compared to models that explicitly account for stochasticity. In fact, a deterministic solution that nicely satisfies an aggregate capacity constraint in order to match supply with demand may be infeasible to implement in practice since the operational dynamics will cause lead times to grow within each time bucket. This results in excessive WIP as well as shortages, and related holding and overdue costs. Not only penalty costs or lost sales are problematic, also the impact of short-term, ad hoc measures that are needed to adapt capacity levels (overtime, alternative routings, outsourcing, etc.) should not be underestimated because of their expensive and disruptive nature. All these negative effects can be avoided when the system is treated as a stochastic network. So there is a need for a decision support tool that accurately determines the lead time effects from workload assignments to every period and to every resource subject to various stochastic elements.

The structure of the paper is as follows. First, the literature related to models that handle uncertainty in aggregate planning decisions is given in Section 2. Section 3 outlines the analytical queueing expressions to estimate all the cost and revenue components in ARP's profit function. This section concludes with a brief presentation of the optimization routine to solve this problem. Results are obtained from a simple, but realistic dataset introduced in Section 4. Managerial insights are derived from a scenario analysis, followed by suggestions concerning the lot sizes to improve the overall system performance. Conclusions are outlined in Section 5.

2 Literature Overview

Based on the distinction by Ho (1989) between environmental uncertainty (i.e. beyond the production process like demand and supply uncertainty), and system uncertainty (i.e. within the production process like process time, quality and breakdowns), Mula et al. (2006) categorize models for production planning under uncertainty into four types: conceptual, analytical, artificial intelligence and simulation. Since the goal of this study is to develop a quantitative model with analytical expressions, the literature overview is focused on the analytical approach. The conceptual models and the artificial intelligence are not covered here because they do not fit the scope of this study. Only the simulation approach is briefly discussed first to show the benefits of an analytical model.

Simulation models evaluate different performance parameters of the production system in an iterative way. Using Monte Carlo simulation to model the uncertainty in costs, capacities, lead times and demand, Thompson and Davis (1990) and Thompson et al. (1993) formulate a linear programming model for the aggregate production planning. Simulation is a flexible approach because details of the processing steps can be modeled, but its vast amount of required runs prevents an efficient performance optimization for our purposes. In addition, simulation models of large systems are difficult to maintain and time-consuming to run and analyze. Scaling is problematic because the number of shop floor decisions tends to grow very fast with the number of products and resources.

In addition to uncertainty, research effort using an *analytical approach* has been invested to handle the disadvantage of using lead times, lot sizes and capacity buffers that are fixed and predetermined in standard APP models. Deterministic techniques that assign capacity to discrete time buckets do not model the operational dynamics of the production system. Changing product mixes, production rates and inventory levels from one period to another in the production plan, may not be executed in reality as easy as the static model suggests. Models that use a fixed lead time like the Materials Requirements Planning approach of Orlicky (1975) and Vollmann et al. (2005), assume infinite capacity, which means that all work can be finished in a fixed amount of time regardless of the workload. However, according to queueing, lead times are affected in a non-linear way by the workload, which is the result of the planning tool assigning jobs to resources. In addition, the outcome of planning models with lot sizes that are fixed during the planning horizon for each product is also not realistic because it does not reflect the dynamics in the system. Queueing models suggest a convex relationship between the lot size and the average lead time and WIP (Karmarkar, 1987; Lambrecht et al., 1998; Lieckens and Vandaele, 2011).

Linear Programming (LP) models that impose a fixed capacity bound are neither a solution because they do not distinguish between finished-goods (FG) and WIP. FG is stock that is built on purpose during a specific time period to be used later, while WIP is built automatically depending on the flows, queues and capacities in the system. LP models are not able to capture these dynamic effects. In addition, their dual price of resource capacity is zero until the capacity constraint is completely saturated, leading to maximal filling up a resource. However, queueing models demonstrate that system performance non-linearly deteriorates as the utilization level approaches 100% (Hopp and Spearman, 2000). Hackman and Leachman (1989) present a generic LP formulation for production planning that minimizes holding and production costs in multi-stage systems, but congestion effects are ignored due to workload independent offsetting. Some authors have developed LP models that include a non-linear term with piecewise linearization to link lead times (Voss and Woodruff, 2003) and costs of holding WIP (Kekre, 1984; Ettl et al., 2000) to the workload. Riano (2003) uses in an iterative LP scheme similar to that of Hung and Leachman (1996) to model random lead times. Ignoring this leads to cost inefficient production plans.

Iterating between LP models that propose production plans and simulation models that evaluate these plans is another recent method Hung and Hou (2001); Hung and Leachman (1996); Byrne and Hossain (2005); Kim and Kim (2001), but simulation problems are experienced with large systems and the convergence properties are not well understood (Irdem et al., 2008).

Apart from simulation models that are computationally intractable and mathematical programming models that cannot account for operational dynamics, analytical clearing functions can be found in the literature as well to capture the nonlinear dynamics with a low computational effort. They use a closed-form expression from queueing theory in order to relate the expected throughput of a capacitated resource to its WIP-level in a planning period, defined by its average utilization (Graves, 1986; Karmarkar, 1989; Srinivasan et al., 1988). Missbauer (2002) optimizes production planning based on clearing functions without modeling product mix effects. Asmundsson et al. (2009) develop an aggregate clearing function for multi-product, multi-stage systems linking expected WIP of all products to the expected output of the resource. They use outer linearization of this nonlinear model to obtain a linear programming formulation and assume no significant effect of lot sizing on system's performance. Their computational experiments indicate that when clearing functions are estimated accurately, these models perform considerably better than LP models with fixed, exogenous lead time estimates. Hopp and Spearman (2000) provide a number of illustrations of clearing functions for a variety of systems, where Srinivasan et al. (1988) derives them for a closed queueing network with a product form solution.

Since this approach is subject to restrictive assumptions, an advanced clearing function to approximate a complex, multi-stage production system as described in Section 1 (multiple products, resources, routings, ...) with significant lot size effects from setup times and general distributions for the stochastic parameters has not been presented yet. In this context, Hahn et al. (2012) propose an iterative framework for APP integrating an Aggregate Stochastic Queueing (ASQ) model that is considerably less restrictive compared to the clearing function approach, while it can be solved to optimality within an acceptable time range. A case-oriented numerical example highlights the improvement potential of this new approach compared to standard APP. We refer to Pahl et al. (2007) for an extensive overview of production planning models with load dependent lead times.

3 Model Formulation

In this section, we develop an analytical expression for the expected profit from aggregate planning decisions in ARP. This model consists of the following elements: multiple products $p \in \{1,...,P\}$, multiple resources $m \in \{1,...,M\}$, multiple product related routings $r \in \{1,...,R_p\}$, multiple operations $o \in \{1,...,O_{pr}\}$ that are product and routing specific and multiple time buckets $t \in \{1,...,T\}$. We use a binary parameter ς_{prom} that is equal to 1 if product p requires machine m at operation o in routing r. Parameters that represent a batch are complemented by a tilde $(\tilde{\ })$. Since the availability is a critical ARP input, it is derived in a separate section, preceded by an overview of the assumptions. Unless otherwise stated, all time parameters are expressed in hours.

3.1 Assumptions

The ARP model is developed to support aggregate planning decisions in a production or supply chain system where the processing is performed according to a first-come, first-serve priority rule. Resources can be either machines or labor, both with single or parallel servers, but we limit the notation to machines m. Forecasts and firm orders drive the demand, which is independent and identically distributed. Similar to their inter-arrival times, the sett-up and process times are assumed to be generally distributed with SCVs to describe the degree of randomness. When a batch is assigned to a routing, the sequence of operations is known prior to its release in the shop (i.e. deterministic routing) and its lot size remains fixed (i.e. no transfer batching). We further assume that set-up times and costs are independent of the product sequence.

The queueing approach extends on Lambrecht et al. (1998). After integrating several queueing approximations found in the literature (Shanthikumar and Buzacott, 1981; Buzacott and Shanthikumar, 1985; Shanthikumar and Sumita, 1988; Bitran and Tirupati, 1988), they estimate the lot size dependent safety lead time with a service level based on the lognormal approximation, followed by lot size optimization to improve the overall due date performance. The underlying queueing model and lot size optimization process are further refined in Lieckens and Vandaele (2011). Other ARP applications can be found in Vandaele et al. (2002, 2003); Nieuwenhuyse and Vandaele (2006); Nieuwenhuyse et al. (2007, 2011).

3.2 Availability

The availability measure A_{mt} is a key ARP component. It is defined as the probability that machine m is available for value-adding activities in period t. Since it represents the percentage of time on a continuous time scale during which machine m is available for processing in period t, we can implement it as a decision variable to determine two elements: the number of shifts per working day $sh_{mt} \in \{1,2,3\}$ (expressed in a number of regular working hours RT_{mt} per shift) and total overtime that should be added (expressed in total absolute working hours OT_{mt}). Furthermore, by using A_{mt} we can transform time dependent queueing parameters into a common time unit, e.g. calendar days when A_{mt} is applied to the number of calendar days CD_t. This makes it easy to interpret delays in the production system, regardless of the scheduled time that may differ for each machine m used in different planning periods t. Sources that influence the available time of machine m in period t include number of working days WD_{mt}, regular working hours per shift RT_{mt}, total overtime OT_{mt}, preventive maintenance PM_{mt} and some period independent characteristics like breakdown (Mean-Time-To-Repair $MTTR_m$ and Mean-Time-To-Failures $MTTF_m$) and an efficiency percentage E_m for any remaining loss in productive time. A time bucket t has a continuous time length of 24 hours/day $\times$ CD_t days/period, which will be denoted by CH_t calendar hours in period t. This results in the following equation for the machine availability during each time bucket t

$$A_{mt} = \frac{(RT_{mt} sh_{mt} WD_{mt} + OT_{mt} - PM_{mt})\, MTTF_m E_m}{CH_t\,(MTTR_m + MTTF_m)} \tag{1}$$

The fraction $MTTF_m/(MTTR_m + MTTF_m)$, which represents the probability that a machine m is not failing, is applied to the net available working time of $RT_{mt} sh_{mt} WD_{mt} + OT_{mt} - PM_{mt}$.

Although part time shifts can be handled, we assume a 3-shift regime for ease of understanding and notation. This involves a lower and upper bound on A_{mt} by replacing sh_{mt} in Equation (1) with 1 and 3 respectively, and $OT_{mt} = 0$. A continuous fraction between these two bounds will be searched for by the algorithm in Section 3.5, in such a way that profit is best. This fraction regulates the capacity level of machine m in period t because the number of shifts (an integer value) and the total number of working hours in overtime (a real value) can be easily obtained by Algorithm 1.

Algorithm 1 Calculate overtime and number of shifts

for $m = 1 \to M$ **do**
  for $t = 1 \to T$ **do**
    $sh_{mt} = \lfloor \frac{1}{RT_{mt}WD_{mt}} \left(\frac{CH_t A_{mt}(MTTR_m + MTTF_m)}{MTTF_m E_m} + PM_{mt} \right) \rfloor$
    if $sh_{mt} \neq 3$ **then**
      $OT_{mt} = \frac{CH_t A_{mt}(MTTR_m + MTTF_m)}{MTTF_m E_m} - (RT_{mt} sh_{mt} WD_{mt} - PM_{mt})$
      if $otc_m OT_{mt} \geq rtc_m RT_{mt} WD_{mt}$ **then**
        $sh_{mt} \leftarrow sh_{mt} + 1$
        $OT_{mt} \leftarrow 0$
        Solve Equation (1)
      end if
    end if
  end for
end for

Given a value for A_{mt}, the first line determines the number of shifts where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. If this is not equal to its upper bound of 3, the total working hours that remain in overtime can be derived. When each hour at machine m is charged rtc_m for regular time and otc_m for overtime, we can calculate the point where it becomes less expensive to add one additional shift instead of overtime. This requires an update of the availability measure. Note that this approach is only valid when the overtime can be assigned to any combination of personnel in such a way that their total working time does not exceed the allowed maximum.

3.3 Arrival Process

Single unit orders for product p with due date in period $t' \in \{1, ..., T\}$ are expected to arrive at a rate $D_{pt'}$ per hour, expressed at a continuous time scale (24 hours per day, $CD_{t'}$ days per period, etc.). The SCV $c^2_{IA_{pt'}}$ describes the variability of their inter-arrival time. Apart from this demand information, we also need to know the arrival rates at resources. These rates are adapted with a scrap fraction ω_{pro} when quality problems occur at operation o. Two distinct, but fractional decision variables (between 0 and 1) are used to obtain the final production rates in each period t

- $\kappa_{ptt'}$ is applied to the demand of product p in period t' to obtain its expected production rate in period t as $\lambda_{pt} = \sum_{t'} \kappa_{ptt'} D_{pt'}$. This volume rate is after the subtraction of scrapped units along the different process steps in all alternative routings (=net output rate after scrap)
- γ_{prt} is applied to this production volume to obtain the expected production rate of product p at routing r in period t as $\lambda_{prt} = \gamma_{prt} \lambda_{pt} / \prod_{o} (1 - \omega_{pro})$. This volume rate still includes units that may be scrapped at later process steps in route r (=brut input rate before scrap)

The condition $\sum_{r} \gamma_{prt} < 1$ involves outsourcing the remaining volume of product p in period t. This means that outsourcing is an alternative routing $R+1$ to resolve capacity problems and that λ_{pt} should be interpreted as the net output rate of the potential production quantity, of which some part may not be manufactured in-house. The fraction γ_{prt}, which assigns some portion of this rate to routing r, must be divided by the multiplication of all net fractions that continue to the next operation in order to compensate for scrap and to have an overall throughput yield equal to total demand. The average arrival rate of product p at operation o in routing r can be written as

$$\begin{aligned} &\lambda_{prot} = \lambda_{prt} && \text{if } o = 1 \\ &\lambda_{prot} = \lambda_{prt} \prod_{o'=1}^{o-1} \left(1 - \omega_{pro'}\right) && \text{if } o > 1 \end{aligned}$$

Each product p is manufactured in batches of Q_p units with zero collection time. Note that this lot size is product specific, but it does not change between periods t and between operations o. The batching process is characterized by average batch arrival rates $\tilde{\lambda}_{pt}$, $\tilde{\lambda}_{prt}$ and $\tilde{\lambda}_{prot}$, and a SCV of batch inter-arrival times $\tilde{c}^2_{IA_{pt'}}$, where each parameter is found by dividing its corresponding parameter value by Q_p. Using the results from Shanthikumar and Buzacott (1981), we approximate the SCV of the batch inter-arrival times at routing r in period t caused by external demand of product p that is due in period t' as

$$\tilde{c}^2_{IA_{prtt'}} \approx \gamma_{prt} \kappa_{ptt'} \tilde{c}^2_{IA_{pt'}} + 1 - \gamma_{prt} \kappa_{ptt'} \tag{2}$$

The queueing network approach requires aggregation of these multi-product, multi-routing batch arrival processes into a single batch arrival process at machine m in period t. It is also characterized by an average aggregate batch arrival rate $\tilde{\lambda}_{mt}$ and a SCV of the aggregate batch inter-arrival times $\tilde{c}^2_{IA_{mt}}$. The aggregate batch arrival rate of product p at machine m in period t can be derived as $\tilde{\lambda}_{pmt} = \sum_r \sum_o \tilde{\lambda}_{prot} \varsigma_{prom}$, leading to $\tilde{\lambda}_{mt} = \sum_p \tilde{\lambda}_{pmt}$. This includes batch arrivals at machine m that are both internal, i.e. coming from another machine, and external, i.e. coming from the customer. The external aggregate batch arrival rate at machine m is derived as $\tilde{\lambda}'_{mt} = \sum_p \sum_r \tilde{\lambda}_{prt} \varsigma_{pr1m}$. Since more information is required on variabilities at upstream production processes to find the internal variability level of product arrivals (see later in Equation (7)), we can only obtain at this moment an approximate SCV of the external aggregate batch inter-arrival times at machines m in period t

$$\tilde{c}'^2_{IA_{mt}} \begin{cases} = \tilde{c}^2_{IA_{pt'}} & \text{if } \sum_p \sum_r \varsigma_{pr1m} = 1,\ \sum_{t'} \kappa_{ptt'} = 1,\ \kappa_{ptt'} = 1,\ \varsigma_{pr1m} = 1 \text{ and } \exists! r : \gamma_{prt} = 1 \\ \approx \frac{1}{3} + \frac{2}{3} \sum_p \sum_r \sum_{t'} \frac{\gamma_{prt} \kappa_{ptt'} \left(D_{pt'} / \prod_o \left(1 - \omega_{pro}\right) \right) \varsigma_{pr1m}}{\sum_p \sum_r \lambda_{prt} \varsigma_{pr1m}} \tilde{c}^2_{IA_{prtt'}} & \text{if otherwise} \end{cases} \tag{3}$$

where the weights $1/3$ and $2/3$ are a specific case of a general approximation found by Albin (1981) and as discussed in Lambrecht et al. (1998).

3.4 Production Process

Each product p that requires processing at operation o along route r has the following characteristics in hours: expected setup time SU_{pro} with SCV $c^2_{SU_{pro}}$ and variance $\sigma^2_{SU_{pro}}$, expected unit processing time PR_{pro} with SCV $c^2_{PR_{pro}}$ and variance $\sigma^2_{PR_{pro}}$, expected unit process rate $\mu_{pro} = 1/PR_{pro}$, and a rework percentage $rwrk_{pro}$. An operation o takes place at machine m by using $\varsigma_{prom} = 1$. This enables machines m to be used more than once along a product's route r. Only one machine type m can be assigned to an operation o, i.e $\sum_m \varsigma_{prom} = 1 \ \forall p, r, o$. Furthermore, each machine can have multiple, parallel servers s_m. The following effective process characteristics are used (Hopp and Spearman, 2000)

- $SUe_{prot} = SU_{pro} / \left(\sum_m A_{mt} \varsigma_{prom} \right)$

- $PRe_{prot} = PR_{pro} / \left((1 - rwrk_{pro}) \sum_m A_{mt} \varsigma_{prom} \right)$

- $\sigma^2_{SUe_{prot}} = c^2_{SU_{pro}} SUe^2_{prot}$

- $\sigma^2_{PRe_{prot}} = \dfrac{2PRe_{prot} \sum_m MTTR_m (1 - A_{mt}) \varsigma_{prom} + c^2_{PR_{pro}} PRe^2_{prot}}{1 - rwrk_{pro}} + \dfrac{rwrk_{pro} PRe^2_{prot}}{(1 - rwrk_{pro})^2}$

These effective measures allow us to express all the delays in the network on a continuous time scale, e.g. calendar days instead of working days. Note that these measures depend on the time bucket t because the available time differs due to different operational time schedules in each period. Total effective production time of a batch with products p at operation o in routing r and period t becomes $TPRe_{prot} = SUe_{prot} + Q_p PRe_{prot}$, which can be used to find the average aggregate batch processing time on machine m in period t as a weighted average

$$\tilde{PR}_{mt} = 1/\tilde{\mu}_{mt} = \sum_p \sum_r \sum_o \frac{\tilde{\lambda}_{prot} \varsigma_{prom}}{\tilde{\lambda}_{mt}} TPRe_{prot} \tag{4}$$

The expression for the corresponding SCV is (Lambrecht et al., 1998)

$$\tilde{c}^2_{TPRe_{mt}} = \sum_p \sum_r \sum_o \frac{\tilde{\lambda}_{prot} \varsigma_{prom}}{\tilde{\lambda}_{mt}} \left[\frac{\sigma^2_{SUe_{prot}} + Q_p \sigma^2_{PRe_{prot}}}{TPRe^2_{prot}} + (TPRe_{prot} \tilde{\mu}_{mt})^2 \right] - 1 \tag{5}$$

The machine utilization ρ_{mt}, which must be lower than 100%, follows as

$$\rho_{mt} = \frac{\tilde{\lambda}_{mt}}{\tilde{\mu}_{mt} s_m} = \sum_p \sum_r \sum_o \tilde{\lambda}_{prot} \varsigma_{prom} TPRe_{prot} / s_m \leq 1 \tag{6}$$

For $\tilde{c}^2_{IA_{mt}}$, the SCV of the aggregate batch inter-arrival times at machine m, the final input parameter of the queueing network, we need to know the variability of the inter-departure times of batches leaving the upstream machine m'. Several approximations for systems with multiple servers exist (see Buzacott and

Shanthikumar (1993)), but we opt for the following linking equation (Hopp and Spearman, 2000)

$$\tilde{c}^2_{ID_{m't}} \approx \left(1-\rho^2_{m't}\right)\left(\tilde{c}^2_{IA_{m't}}-1\right)+1+\frac{\rho^2_{m't}}{\sqrt{s_{m'}}}\left(\tilde{c}^2_{TPRe_{m't}}-1\right) \tag{7}$$

It is affected by various operational decisions that have an impact on the utilization level (Equation (6)) and the variability levels (Equations (5) and (8)) at machine m'. These levels depend on the total load that results from periodic and routing assignment decisions as well as lot size decisions. From here, some fraction of the product flow is sent to a downstream machine m that is equal to

$$f_{m'mt} = \sum_p \sum_r \sum_o^{O_p-1} \tilde{\lambda}_{prot} \varsigma_{prom'} \varsigma_{pr(o+1)m} / \tilde{\lambda}_{m't}$$

We can use an equation similar to (2) if we want to split the variability of the departing flow into distinct components. This results in the following value for the SCV of the aggregate batch inter-arrival time at a downstream machine m for products coming from an upstream machine m'

$$\tilde{c}^2_{IA_{m'mt}} = f_{m'mt}\tilde{c}^2_{ID_{m't}} + (1-f_{m'mt})$$

Using this equation in the weighted average expression for the required SCV

$$\tilde{c}^2_{IA_{mt}} = \sum_{m'}\left(\frac{\tilde{\lambda}_{m't}}{\tilde{\lambda}_{mt}} f_{m'mt}\right)\tilde{c}^2_{IA_{m'mt}} + \frac{\tilde{\lambda}'_{mt}}{\tilde{\lambda}_{mt}}\tilde{c}'^2_{IA_{mt}} \tag{8}$$

in combination with Equation (7) leads to M linear equations with M unknown parameters $\tilde{c}^2_{IA_{mt}}$. They are obtained as in Lambrecht et al. (1998). We continue to use the approximation from Whitt (1993) for the GI/G/m queueing model to estimate the expected total lead time of product p in routing r in period t by using Equations (4), (5), (6) and (8)

$$EW_{prt} \approx \sum_o \left(\sum_m EWQ_{mt}\varsigma_{prom} + SUe_{prot} + +Q_p PRe_{prot}\right) \tag{9}$$

with

$$EWQ_{mt} \approx \phi\left(\rho_{mt}, \tilde{c}^2_{IA_{mt}}, \tilde{c}^2_{TPRe_{mt}}, s_m\right)\left(\frac{\tilde{c}^2_{IA_{mt}} + \tilde{c}^2_{TPRe_{mt}}}{2}\right)\left(\frac{\rho_{mt}^{\sqrt{2(s_m+1)}-1}}{s_m(1-\rho_{mt})}\right)\tilde{PR}_{mt}$$

Whitt (1993) also provides an approximation for the period dependent variance of the waiting time at machine m, which is equal to VWQ_{mt}. The variance of the total lead time of product p in routing r in period t becomes

$$VW_{prt} \approx \sum_o \left(\sum_m VWQ_{mt}\varsigma_{prom} + \sigma^2_{SUe_{prot}} + Q_p\sigma^2_{PRe_{prot}}\right)$$

When the outsourcing option is chosen (i.e. alternative routing $R+1$), an average and variance of this delay are assumed to be given in each time bucket. These relationships enable us to calculate an expected aggregate lead time as a weighted average

$$EW_{pt} \approx \sum_{r=1}^{(R+1)} \pi_{prt} EW_{prt} \tag{10}$$

with corresponding variance

$$VW_{pt} \approx \sum_{r=1}^{(R+1)} \pi_{prt}\left(EW^2_{prt} + VW_{prt}\left(\frac{EW_{pt}}{EW_{prt}}\right)^2\right) - EW^2_{pt} \tag{11}$$

and where the weights are given by (explanation at the end of this section)

$$\pi_{p(r\neq R+1)t} = \frac{\gamma_{prt}\lambda_{pt}}{\sum_r \gamma_{prt}\lambda_{pt} + MAX\left[\left(1-\sum_r \gamma_{prt}\right)\lambda_{pt} - \sum_{t'>t}\kappa_{ptt'}D_{pt'}, 0\right]} \tag{12}$$

$$\pi_{p(R+1)t} = \frac{MAX\left[\left(1-\sum_r \gamma_{prt}\right)\lambda_{pt} - \sum_{t'>t}\kappa_{ptt'}D_{pt'}, 0\right]}{\sum_r \gamma_{prt}\lambda_{pt} + MAX\left[\left(1-\sum_r \gamma_{prt}\right)\lambda_{pt} - \sum_{t'>t}\kappa_{ptt'}D_{pt'}, 0\right]} \tag{13}$$

Suppose that the earliest due date of orders asked for in period t is after DW_p hours at a continuous time scale and that the gap between release and due date of orders that are both scheduled and due in the current time bucket is always DW_p hours as well. This means that the time between zero and DW_p can be used as a phasing-out period for units from the previous period that are not finished on time, and as a phasing-in period to start working on a first-come-first-serve basis on units that are scheduled in the current period and due DW_p hours later. Time buckets at the intermediate S&OP level can be considered long enough to satisfy steady state conditions. The time delay DW_p is composed of two components: an average delay and a safety time. The question is which service level SL_{pt} is to be expected? In other words, what is the desired probability that the due date will be met? We follow an approach similar to Lambrecht et al. (1998) by postulating a lognormal distribution for the aggregate lead time. After deriving its shape and scale parameters, the z_{pt}-value from the standard normal distribution with cumulative probability SL_{pt} follows as

$$z_{pt} \approx \frac{ln\left(DW_p\right) - ln\left(\frac{EW_{pt}}{\sqrt{VW_{pt}/EW_{pt}^2+1}}\right)}{\sqrt{ln\left(VW_{pt}/EW_{pt}^2+1\right)}} \tag{14}$$

The lead time information EW_{pt} and VW_{pt} is only used in this expression for the service level SL_{pt} and this measure will only be used to calculate the backorder costs for units that are demanded and produced in the current period t (see Equation 20 in Section 3.5). As a result, it would be unfair to include the lead time impact of units that are outsourced in period t but demanded in later periods $t' > t$ into the total lead time calculations in Equations 10 and 11. To this end, we assume that

- outsourced units in period t will never be used to satisfy demand in earlier periods $t' < t$
- outsourced units in period t are primarily used to satisfy demand in later periods $t' > t$ when there are assignments from this future demand to the current period t (i.e. when $\kappa_{ptt'} > 0$)
- the outsourced units that remain in period t are further used to satisfy demand in t

This means that outsourcing is not used to fulfill backorders, but to build-up inventory either for future or current demand satisfaction. The total demand volume of future periods $t' > t$ assigned to period t is equal to $\sum\limits_{t'>t} \kappa_{ptt'} D_{pt'}$. When this volume is subtracted from the outsourcing quantity $\left(1 - \sum\limits_{r} \gamma_{prt}\right) \lambda_{pt}$, the leftover is the relevant outsourced volume that is allowed to affect the weights in Equations 12 and 13, where the maximum function ensures non-negative values. In this way, the service level SL_{pt} in the current period t only depends on in-house production in the current period t (Equation 12) and on units outsourced and demanded in the current period t (Equation 13).

3.5 Objective Optimization

Profit as the objective function includes all cost components that are relevant during the aggregate planning horizon, i.e. WIP, FG, setups, labor, outsourcing, backorders, disposals and other variable operating issues. Revenue is generated when a product p is sold at a net unit selling price sp_p. Although more sophisticated models exist where price is a function of time, demand volume and/or delivery guarantee (see Upasani and Uzsoy (2008) for an extensive overview), we use a fixed value for this parameter. Financial implications due to delays and advances of the cash flow over time are not included, but can be easily incorporated. Total demand in the planning horizon must be satisfied under the assumption that the customer pays the total amount in period t where he wants the product p to be delivered. As a result, total expected revenue in period t is equal to

$$REV_t = \sum_p CH_t D_{pt} sp_p \tag{15}$$

WIP are units that naturally arise at the inbound of operation o in routing r. On average, it costs hc_{pro} to keep one unit of product p in the buffer during one time bucket. This unit cost is assumed to be independent of period t. Since the planning periods are long enough to satisfy Little's Law, total expected WIP costs in period t become

$$WIPC_t \approx \sum_p \sum_r \sum_o \lambda_{prot} hc_{pro} \sum_m EW_{mt} \varsigma_{prom} \tag{16}$$

Each time a batch enters an operation o in routing r, a setup time and associated cost suc_{pro} are charged. Since the total number of batches in each time bucket is known, total expected setup costs in period t

become

$$SUC_t = CH_t \sum_p \sum_r \sum_o \tilde{\lambda}_{prot} suc_{pro} \tag{17}$$

Products that are not manufactured in-house are supplied by a subcontractor at a time independent unit cost of osc_p, leading to an expected outsourcing cost in period t of

$$OSC_t = CH_t \sum_p \lambda_{pt} \left(1 - \sum_r \gamma_{prt}\right) osc_p \tag{18}$$

Labor is paid during regular time and overtime at a rate rtc_m and otc_m per hour. The required number of operators is not directly modeled and is assumed to be fixed, but their financial impact can be incorporated into these rates. Note that number of shifts sh_{mt} and overtime OT_{mt} in each period t are determined in Algorithm 1. We have resource dependent tariffs because we want to distinguish different levels of operational complexity. Total expected labor costs in period t become

$$LC_t = \sum_m (WD_{mt} sh_{mt} RT_{mt} rtc_m + OT_{mt} otc_m) \tag{19}$$

Backorders occur when production is delivered after the due date. A product specific cost boc_p is then charged for each period being too late. Total expected backorder costs in period $t > 1$ become

$$BOC_t \approx \sum_p \left(\sum_{t'=1}^{t-1} CH_{t'} D_{pt'} \kappa_{ptt'} boc_p \left(t - t'\right) + \sum_t CH_t D_{pt} \kappa_{ptt} \left(1 - SL_{pt}\right) boc_p \right) \tag{20}$$

The first part is for product volume that is produced in a time bucket t later than each time bucket t' where a demand occurs ($t > t'$). This volume is equal to $CH_{t'} D_{pt'} \kappa_{ptt'}$. The backorder cost boc_p applies for each delayed period. For example, if we produce in period $t = 4$ for demand in period $t' = 1$, the unit backorder cost is taken three times ($t - t' = 3$). The second part is for product volume that is produced in the same time bucket t as the demand occurrence, with a probability of being too late equal to $(1 - SL_{pt})$. When backorder costs would be charged for demand volume that is transferred to a period beyond the planning horizon, the algorithm would favor this kind of solutions because of a smaller total cost. The reason is that this strategy avoids any other cost types except the backorder cost. Therefore we do not allow products to be assigned to periods $T+x$ with $x>0$. As a result, all demand in each period t' will be produced within the planning horizon.

FG inventories occur when production is finished before the demand due date. A product specific cost fgc_p is charged for each period being too early. Total average FG costs in period t become

$$FGC_t = \sum_p \sum_{t'=t+1}^{T} CH_{t'} D_{pt'} \kappa_{ptt'} fgc_p \left(t'-t\right) \tag{21}$$

Demand that occurs in a time bucket t' later than the current production slot t must be held in stock during $t'-t$ periods at a holding cost of fgc_p per period.

Scrap is disposed of at a unit cost dc_{pro}, leading to an expected disposal cost in period t of

$$DC_t = \sum_p \sum_r \sum_o CH_t \lambda_{prot} \omega_{pro} dc_{pro} \tag{22}$$

All other operational costs that vary with the produced volume but not included in costs modeled so far (raw material, supplies, transportation, electricity, etc.) are bundled into a variable unit cost vc_{pro}. They can be used to distinguish between alternatives, where the main route is usually less expensive than a secondary, emergency route. Total expected variable costs in period t become

$$VC_t = \sum_p \sum_r \sum_o CH_t \lambda_{prot} vc_{pro} \tag{23}$$

Equations (15)-(23) can be aggregated to obtain the overall profit function

$$\pi \approx \sum_t \pi_t \approx \sum_t \left[REV_t - \left(WIPC_t + SUC_t + OSC_t + LC_t + BOC_t + FGC_t + DC_t + VC_t\right)\right] \tag{24}$$

subject to the constraints

$$\sum_t \kappa_{ptt'} \leq 1 \qquad \forall p, t'$$

$$\sum_r \gamma_{prt} \leq 1 \qquad \forall p, t$$

$$\rho_{mt} \leq 1 \qquad \forall m, t$$

$$\frac{(1 \times RT_{mt} WD_{mt} - PM_{mt}) MTTF_m E_m}{CH_t (MTTR_m + MTTF_m)} \leq A_{mt} \leq \frac{(3 \times RT_{mt} WD_{mt} - PM_{mt}) MTTF_m E_m}{CH_t (MTTR_m + MTTF_m)} \qquad \forall m, t$$

$$0 \leq \kappa_{ptt'} \leq 1 \qquad \forall p, t, t'$$

$$\gamma_{prt}^{min} \leq \gamma_{prt} \leq \gamma_{prt}^{max} \qquad \forall p, r, t$$

$$Q_p^{min} \leq Q_p \leq Q_p^{max} \qquad \forall p$$

The first two constraints ensure that total production quantity in the planning horizon to satisfy demand in some period does not exceed this demand and that the product flow assigned to the routings in each period does not exceed the scheduled production volume. Outsourcing takes the remaining units. The third constraint controls the capacity feasibility of the aggregate plan. The other constraints impose lower and upper bounds on the decision variables.

Before we describe the decision variables, we recall the goal of this paper: finding a good balance between production volume and resource supply in each time bucket to maximize overall profit in the planning horizon at the S&OP level. The ARP model can contribute to the effectiveness and efficiency of the S&OP meetings by searching for optimal values of the decision variables. The first set consists of the variables $\kappa_{ptt'}$ that determine the product-mix to be released in the shop floor in each period. The second set of variables deals with the supply side by making a trade-off between various alternative capacity decisions. Here, variables γ_{prt} decide on the flow assignment to each routing where outsourcing supplies the units that are off-loaded to an external provider, while variables A_{mt} decide on the number of shifts and overtime that must be dedicated to each resource in each period. The third set relates to the lot size variables Q_p that can further improve the performance of the entire system through lead time reduction. We have found evidence that the convex relationship between lot size and lead time still holds in complex production systems with multiple products and multiple resources (Lieckens and Vandaele, 2011). On the one hand, lot sizes that are too small result in high values for the expected lead time in Equation (9) because the first term for the queue time increases in a non-linear way due to an enhanced utilization level caused by setting-up resources more frequently (saturation effect). On the other hand, lot sizes that are too large also result in high values for the expected lead time in Equation (9) because the second term for the total process time linearly increases with the lot size quantity (batching effect). In between, an optimal lot size exists that minimizes this trade-off. Since lead times, and thus also the service levels in Equation (14), clearly depend on the lot size, the profit function depends on it as well through the costs for WIP and backorders. Note that the number of machines m is not a decision variable because it is considered to be a long term strategic decision that is not relevant at this planning level.

We summarize the characteristics that make it difficult to solve the ARP model:

- Decision variables $\kappa_{ptt'}$ are fractions that drive the combinatorial nature of production volume and product mix in period t caused by demand in period t'
- Decision variables γ_{prt} are fractions that further extend the search space with possible assignments of products to routings
- Decision variables A_{mt} are fractions that drive shift and overtime decisions
- Decision variables Q_p are discrete values that drive trade-offs between saturation and batching effects
- Expected waiting times are non-linearly dependent on the utilization level ρ_{mt}
- Lead times and service levels create a non-linear objective (Equation (24))
- The M constraints in Equation (6) are non-linearly dependent on Q_p.
- The variability coefficients depend on all the decision variables in a non-linear way
- There are multiple conditional relationships in $\phi(\rho_m, \tilde{c}^2_{IA_m}, \tilde{c}^2_m, s_m)$ and VWQ_{mt}

The model can be classified as a mixed integer non-linear programming problem (MINLP). In these problems, two difficulties are combined: the combinatorial nature of discrete programs and solving non-linear programs. The computational complexity not only grows exponentially with the number of discrete variables and the number of decisions within each discrete variable, but also with the number of variables creating non-linearities into the model. Storn and Price (1997) have listed nice characteristics of the

Differential Evolution (DE) algorithm to solve MINLP problems. Since we have shown in Lieckens and Vandaele (2011) that DE also performs very well for lot size decisions in an advanced resource planning environment with a single period, we continue to use it as an appropriate solution method for this stochastic multi-period planning problem.

Briefly summarized, the first step of DE is the creation of an initial population of different elements. Each element contains a value for each decision variable ($\kappa_{ptt'}$, γ_{prt}, A_{mt} and Q_p), randomly selected between their lower and upper bounds. Feasible solutions are initially not guaranteed, but by taking the difference of randomly sampled population elements in the mutation and recombination process in combination with a constraint handling procedure, feasible and better performing children are obtained in the population as the number of generations grows. Different mutation schemes can be selected, but they all reflect information of the objective function being optimized. Instead of using only local information for each population element, DE mutates all elements with the same universal distribution. In this way, the whole search space can be covered and a global optimum can be found. Another advantage of DE for mixed integer problems is that discrete variables (Q_p) are treated as continuous values to create subsequent children (this maintains the diversity of the population), while they are rounded at the moment of objective function evaluation. This avoids sub-optimal results because only feasible solutions give feedback to the optimization process. We refer to Lieckens and Vandaele (2007) and Lieckens and Vandaele (2012) for more details on how the algorithm is applied. In order to significantly enhance computing efficiency, an intelligent initialization process that makes use of our specific problem structure is presented in the next section.

4 Model Results

4.1 Base Case

The ARP model is applied to a simple example with two products P1 and P2, each having two routings R1 and R2 with product specific operations: P1 has three operations in both routings, P2 has two operations in both routings. Figure 1 visualizes the process layout. The input data can be found in A (Table 5 for product data, Table 6 for operations data and Table 7 for resource data). In order to test the optimization capabilities and to derive useful insights from a scenario analysis, we have constructed a dataset with small differences between products and routings. The variability level is one of the distinguishing factors between the routings: R1 has always one process step (i.e. second operation O2 for P1 at M2 and first operation O1 for P2 at M4) that is highly variable (SCV=1.5) compared to a low variability level of the corresponding process step in R2 (SCV=0.5). Since these process steps in R1 are also slightly more expensive, we may expect solutions with higher release fractions towards R2, the cheaper route with lower variability. This will be tested below. For the output results we refer to B.

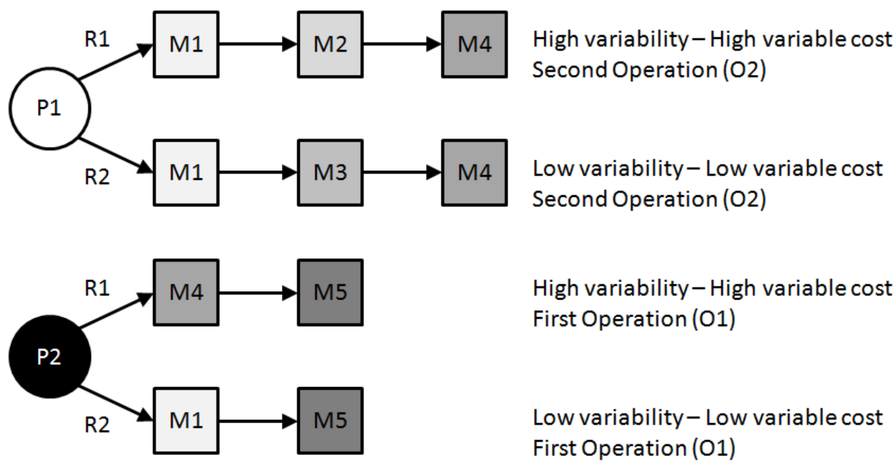

Figure 1: Process layout.

Each operation is performed on a specific resource, chosen from a set that contains five resources with similar characteristics. They have at least one shift, maximum three shifts. According to Equation (1), A_{mt} must be between 23.66% and 70.97%. There are three time buckets in the planning horizon ($T = 3$). We do not allow production in the past ($t \geq 1$) or beyond the planning horizon ($t \leq 3$). This implies that all demand in each period t must be satisfied during the planning horizon, either by in-house production or by outsourcing. We further limit the combinatorial possibilities by imposing the constraint that the demand volume in period t can be produced in the same period t or in periods ± 1 if such a period falls within the planning horizon. For instance, the demand in T1 can only be produced in T1 and T2. This example is rather small to keep the scenario analysis comprehensible and to reach an optimal solution within an acceptable time limit. On average, it takes 3.2 minutes on a 2GHz Intel Core i7 with 8GB 1,333MHz DDR3 for one scenario. Scalability for implementation in practice should be tested further, but experience with the dataset in this study learns us that when near optimal solutions are more than satisfactory, computation time is just a fraction of the total time that is required to find the best solution.

Initialization of the population is critical with respect to computation time, especially when there is a mix of decision variables. Given a specific candidate solution, we do not set initial values for all its parameters in a random and independent way, but use relevant information from parameter values already determined to randomly generate values for remaining parameters. This intelligent initialization process avoids many useless generations because this approach ensures that all candidate solutions are feasible at the beginning of the mutation and recombination process. More specifically, the fractions that assign the demand to a time bucket are generated first. When multiple time buckets are possible, the fraction for the next time bucket is randomly chosen between zero and 100% minus the sum of the fractions over all periods generated so far. The last time bucket receives the remaining fraction because all demand must be satisfied in the planning horizon. For the release fractions that assign volume to routings, a similar approach is taken, except that the last routing R also receives a random, but feasible value. The remaining fraction is automatically interpreted and handled as outsourcing. After these two steps, the magnitude of the production flow at each resource is known. This implies that values for the availability A_{mt} can also be generated in a smart way. As long as some resource m in period t is over-utilized, a new random value for A_{mt} is chosen between its current level and 70.97%. If it approaches this upper bound with some precision ε for at least one resource, the solution is considered to be infeasible, and the process of selecting fractions is started again. This process for the initialization of the population, the first step of DE, is repeated until all resources have a utilization lower than 100%.

The results for the base case are displayed in Table 8. Because of some minor differences between results, we opt not to round volume quantities. Almost 6.5 units of demand for P1 in T2 are produced in T1, creating finished good costs in T1. This additional production volume justifies the application of an additional shift at M1 in T1. This shows that the optimization procedure precisely balances the trade-off between these costs for finished goods and labor. P1 is solely assigned to R2, which is logic from a production perspective (low-cost, low-variability). P2 has a preference for R1, which is more expensive and more variable. This counterintuitive result is explained by the product-process-cost structure. Even when taking into account its lower demand volume, P1 consumes more time at the shared resource M1.

Table 1: Scenario analysis

Scenario	Parameter	Code	Level 1	Level 2	Scenario	Parameter	Code	Level 1	Level 2
1	$c^2_{IA_{pt'}}$	P1	0.25	1.75	7	SU_{pro}	P1-R2-O2	1.25	
2	$8 \times rtc_m$	M5	75		8	hc_{pro}	P1-R2-O2	0	8
3	E_m	M5	0.8		9	$c^2_{SU_{pro}}$	P1-R1-O2	0	1
4	DW_p	P1	100	300	10	$c^2_{PR_{pro}}$	P1-R1-O2	0	1
5	fgc_p	P1	6	18	11	$EW_{p(R+1)}$	P1	40	60
6	boc_p	P1	60	180					

Therefore, it has a preference for the cheaper, more stable route R2. Nevertheless, assigning P2 to this most favorable route R2 as well, would require such an increase in required time at M1 that the savings from R2 cannot compensate the additional capacity costs. It is better to switch to R1 where M4 is used instead because this strategy balances the load over the available resources better. Because of the additional shift at M1 in T1, a significant proportion of P2 can be assigned to R2 because it is cheaper to utilize the capacity jump at M1 than charging overtime at M4. In this base case setting, outsourcing is never used as an alternative capacity resource.

4.2 Scenarios

The next step in the analysis is to test the impact of several parameters on the planning solution, while keeping the lot sizes Q_p fixed. Some interesting insights can be derived from the results that correspond to the scenarios in Table 1. A summary of the main results can be found in Table 2.

The results for Scenario 1, where the variability of the demand interval is increased for P1, are displayed in Table 9. The negative impact of variability on profit and lead times is handled by several planning strategies in order to provide a safety cushion: more units of P1 demand in T2 being produced in T1, more overtime at all machines and a stronger preference for R1 to produce P2. Clearly, performance regarding lead times and overall profit deteriorates.

The results for Scenario 2, where the labor cost for regular time is reduced at M5, are displayed in Table 10. Instead of moving P1 demand earlier in production planning, this is now performed for P2: almost 70 units of its demand in T2 are produced in T1. The increased flow that results in T1 is handled by an ad-

Table 2: Summary results of the scenario analysis

Scenario	Parameter	Key Impact	Profit	Lead Time
1	Variability demand interval ↑	Δ production plan, overtime ↑	↓	↑
2	Labor cost regular time ↓	Δ production plan, shifts ↑	↑	P1↑, P2↓
3	Resource efficiency ↓	Δ production plan, shifts ↑	↓	↑
4	Delivery time guarantee ↑	chase strategy ↑, Δ routing, capacity ↓	↓	more variable
	Delivery time guarantee ↓	finished goods inventory ↑, outsourcing ↑, capacity ↑	↑	↑
5	Holding cost finished goods ↑	Δ production plan, overtime ↑	↓	↓
6	Backorder cost ↓	chase strategy ↑, shifts ↓, overtime ↑	↑	↑
7	Setup time ↑	Δ production plan, overtime ↑	↓	↑
8	WIP cost ↑	Δ production plan, overtime ↑	↓	↓
9	SCV setup time ↑	Δ production plan, overtime ↓	↓	↑
10	SCV process time ↑	Δ production plan, overtime ↑	↓	↓
11	Outsourcing lead time ↑	in-house ↑, Δ routing, overtime ↑	↓	↑

ditional shift at M5, generally avoiding expensive overtime. The routing assignments are approximately the same and overall profit is better. P2 benefits from a lead time reduction at the expense of an extended lead time for P1, a product that is not using M5.

The results for Scenario 3, where the efficiency of M5 is reduced, are displayed in Table 11. Although more pronounced, the impact on production quantities and available working time is similar to the scenario with lower labor costs, while profit is negatively affected.

The results for Scenario 4, where the delivery time guarantee is changed for P1, are displayed in Table 12. This is the time between release and promised due date for products being produced in the same period as their demand occurs. It is also referred to as lead time offsetting. When this time is longer, all demand is produced in the time bucket where it is due (i.e. chase strategy), total available working time is lower (regular time plus overtime) and R2 and R1 are exclusively used for P1 and P2 respectively. This results in an enhanced total profit and lead times that are more variable for P1 and P2. On the other hand, when the lead time offsetting is lower, more finished good inventories of P1 are created because all its demand is produced one period earlier, except the demand in T1 due to the time boundaries in this problem. This is partly solved by using outsourcing as an alternative supply source for P1 in T1: total production quantity of 277 units is reduced by almost 64 outsourced units. This setting requires more shifts but less overtime. We also observe that R1 for P1 and R2 for P2 are activated to a larger extent when compared to other scenarios. All these measures are taken to limit the number of backorders. Nevertheless, profit suffers.

The results for Scenario 5, where the cost to hold finished goods is changed for P1, are displayed in Table 13. When this cost is low, more units of P1 demand in T2 are transferred to production slot T1. This strategy can save on total overtime, leading to a better financial performance. It also justifies longer product lead times. In addition, a lower fraction of P2 is sent to R2 in T1. The opposite is true for a high value for finished goods.

The results for Scenario 6, where the backorder cost for P1 is changed, are displayed in Table 14. High or low values both lead to a chase strategy. When the backorder cost is low, we see a reduction in the number of shifts being compensated by more overtime. It is clearly not a problem that lead times become long and that there are much more backorders and associated total cost (despite a lower unit cost) because the savings in labor costs contribute to a higher profit. Note that no units of P2 are sent to R2 in T1. When the backorder cost is high, the number of shifts remains the same, but more overtime avoids expensive backorders by lowering the product lead times.

The results for Scenario 7, where the setup time for P1 is increased at O2 in R2, are displayed in Table 15. Clearly, less P1 demand is moved from T2 to T1, more overtime is applied and the routing fractions are changed in T1. P1 is also processed at R1 (more variable, more costly but faster setups), while P2 is more in favor of the cheaper, more stable R2 to compensate for the profit loss. This also improves P2's lead time, while it obviously takes longer to deliver P1. For this scenario, we want to investigate why $\gamma_{2r1} = \{0.31, 0.69\}$ in period T1 is better than for instance alternative HighRoute1 with $\gamma_{2r1} = \{0.32, 0.68\}$. In Table 16, we only visualize its resource and product performance measures relative to the results found in Scenario 7. Table 6 and Figure 1 are useful here because it is important to know the operations-resource relationships.

In Scenario 7 we have $\rho_1 > \rho_4$, but the gap becomes smaller in the proposed alternative solution (i.e. more balanced). The lower utilization of M1 leads to a reduction in the expected and variance of its waiting time. The SCV of its external arrivals, which is equivalent to its aggregate SCV of arrivals because M1 is always the first machine in a sequence of operations, is higher because of the lower weight of P2 assigned to R2, a route that has a relative small variability term in Equation 3 when compared to R1 and R2 for P1 (where M1 is used as a first step as well). This is partly due to a larger lot size for P2 that creates a smaller value for Equation 2. The negative queueing effect of this increased SCV of arrivals is opposed by a reduction in the SCV of the process time. This reduction is caused by assigning more weight to P1 and its higher total production time at M1 as described by $TPRe_{prot}$, leading to a higher value for its overall process time in Equation 4. This higher value is responsible for the reduction of $\tilde{c}^2_{TPRe_{m1}}$ because it is in the denominator of Equation 5. The opposite relationships are observed for parameters and measures of M4. Furthermore, the inbound variability levels at M2, M3 and M5 are higher through the linking equations: M1, which feeds these machines, has a lower utilization and a higher SCV of arrivals, resulting into a higher SCV of departures. This explains their deteriorated queueing measures. It proofs that a variability increase at the beginning of the production line negatively affects downstream process steps.

When we look at the details of the product-routing measures in Table 17, we see that EW_{221} is reduced by a lower utilization at M1 and that EW_{211} is increased by a higher utilization at M4. Due to the non-linear queueing relationships, we know that a decrease in the utilization of a highly used machine like M1 has relatively more positive impact on the queueing measures (average and variance) of P2 in R2 than the negative impact from an increase in the utilization of a lowly used machine M4 in R1 for P2. The net result is a better service level for P2. A similar reasoning explains the better service performance for P1, despite of a higher value of EW_{121}. This results from the combined effect of longer delays at M3 and M4 that slightly outweigh the benefit from a shorter delay at M1. However, its major variability reduction, which makes it a more stable process, supports the better overall measures for P1. Better service delivery is translated into a cost reduction for WIP (-0.16%) and backorders (-0.51%). Nevertheless, these benefits are not able to counter the more expensive variable production costs in R1, the route that is used more intense in this HighRoute1 alternative. The end result is a lower profit than in Scenario 7 (-0.0004%), which demonstrates that balancing the utilization (cfr. smaller gap between ρ_1 and ρ_4) is not beneficial. A better profit is neither found in the LowRoute1 alternative with $\gamma_{2r1} = \{0.30, 0.70\}$. In this case, all the effects are mirrored. This shows that the trade-offs between all the cost and queueing relationships in the model are precisely balanced by the DE search algorithm.

The results for Scenario 8, where the WIP cost for P1 is changed at O2 in R2, are displayed in Table 18. As WIP becomes more expensive, there is a slight decrease in the volume of P1 transferred from T2 to T1. In combination with more overtime and an updated release fraction for the routings of P2 in T1, the lead times of both P1 and P2 can be reduced in T1 and T2. Even though the effect is marginal with this minor parameter change, it shows the cost trade-offs between the three buffers (inventory, capacity and time). In order to limit the overall payment for overtime, and consequently also the profit reduction, the model decides to reduce this capacity source in T3, leading to longer delays in that period.

In order to show what kind of impact the variability of setup and production processes can have, we change values for $c^2_{SU_{112}}$ (Scenario 9) and $c^2_{PR_{112}}$ (Scenario 10) in the base case with $suc_{122} = 90$ (see Tables 19 and 20 respectively). This means that the setup cost for a similar process of P1 in R2 is more expensive than the base case. This makes R2 less favorable for P1, which can be observed in Table 19: the release fraction to R2 is lower in T1 when compared to Table 8 where it is 100%. Nevertheless, this fraction increases when the setup process is more variable in R1. The fraction of P1 demand in T2, but produced in T1 also slightly increases. It seems that in this case, longer lead times and less safety capacity (less overtime) are preferred. The net impact on total profit is a marginal decrease. The SCV effects of the production time are more pronounced. When there is no variability, the preference of P1 for R2 disappears in favor of R1. As the variability in R1 increases, the workload in R2 becomes more intense. In contrast to the scenario for setup SCV, the fraction of P1 demand in T2, but produced in T1 decreases. It also becomes worthwhile to invest in more overtime. The positive effects from the reduction in P1's lead time, which results from this additional capacity, is able to limit the profit loss caused by the uncertainty.

The impact of delays in the outsourcing process is analyzed in Scenario 11 by changing values for $EW_{1(2+1)}$ in the base case with $osc_1 = 110$ (see Tables 21). For this setting, there is a positive outsourced volume for P1 in all periods. As expected, in-house production of P1 and required overtime are raised as it takes longer to wait for outsourcing supply. The release fraction of P1 in R2 is also higher. Since DE finds solutions where the outsourced volume is increased as long as it is less costly than in-house production, it is logic that when outsourcing is less efficient due to longer delays, overall profit is worse. The lead times, also for P2, are negatively affected as well.

4.3 ARP versus Standard APP

Another question is to what extent the planning solution proposed by the ARP model differs from more traditional aggregate planning models, or standard APP. To answer this question, we have optimized the ARP model without the disturbing queueing effects caused by the stochastic behavior in the production system (all SCV's equal to zero). In this way, we obtain a more conservative solution that is generated by a deterministic model. In order to provide some safety capacity, we impose an arbitrary and static upper bound on the utilization level of 90% for all resources in all periods, a typical approach in current business practices. The outcome is not directly comparable with the stochastic ARP solution because in the deterministic version of the model some cost components are either missing (e.g. WIP) or calculated differently (e.g. backorder costs due to a service level of 100% in the deterministic model). Therefore, the profit of the solution obtained by the APP model is recalculated with the ARP model in order to get an idea about what the real profit and service level of the deterministic solution would have been in a stochastic environment.

By comparing this revised profit and service value with the optimal ARP profit, we can conclude that the ARP model always significantly outperforms the APP model in terms of profit and delivery performance. See Table 3 for a comparison of overall profit for the scenarios in Section 4.2. More importantly, the production plans are also always completely different. We illustrate this with the deterministic solution of the base case in Table 22. When compared to the results in Table 8, its profit, which is generated according to the same objective function as the base case (Equation 24), is significantly worse, while the production control parameters are different. There is some outsourcing of P1 in T1, no stock of P1 is built-up in T1, the release fraction of P2 to R1 is always smaller than 100% and most importantly, M1 requires one shift less and almost no overtime is needed. Many machines are loaded up to the maximum utilization of 90% (bold numbers), a procedure commonly observed in traditional planning systems. The service level performance, especially of P1, significantly suffers from this loading method. These findings are an indication not only of DE's optimization capability, but also of ARP's strong performance.

Table 3: *Comparison of profit with standard APP for the scenario analysis*

	Level	1			2		
Scenario	Parameter	ARP	APP	Δ	ARP	APP	Δ
1	$c^2_{IA_{pt'}}$	40,573	9,459	329%	40,124	7,879	409%
2	$8 \times rtc_m$	45,574	13,604	235%			
3	E_m	37,146	2,342	1,486%			
4	DW_p	35,705	-5,832	712%	43,550	19,885	119%
5	fgc_p	40,420	8,654	367%	40,341	8,654	366%
6	boc_p	41,678	22,665	84%	39,783	-5,356	843%
7	SU_{pro}	40,038	6,944	477%			
8	hc_{pro}	40,360	8,660	366%	40,334	8,648	366%
9	$c^2_{SU_{pro}}$	39,912	8,120	392%	39,911	8,102	393%
10	$c^2_{PR_{pro}}$	40,002	8,346	379%	39,914	8,174	388%
11	$EW_{p(R+1)}$	41,268	8,929	362%	40,490	8,776	361%

Planning solutions generated by a deterministic approach may not be feasible, or are at least inferior when the production system is stochastic and when service levels are important. Overall profit and delivery performance can always be improved by proposing another structure of the production plan with respect to release fractions, lot sizing (see next), labor time, inventory, backorders, outsourcing, etc.

4.4 Lot Size Optimization

Finally, the impact of the lot size is also investigated. Two approaches have been followed. In the iterative optimization mode, several cycles of production plan and lot size optimization, where one of each is kept fixed in an alternating way, are performed until further profit improvements become marginal (or usually even become worse again). In the simultaneous optimization mode, the lot size decision is just another decision variable in the ARP model that is concurrently considered with the other variables. In order to get an idea about the sensitivity of the product dependent lot size, two versions are compared for this mode: one with product lot sizes that remain the same in the planning horizon T (i.e. static version) and another one with period dependent lot sizes for each product (dynamic version). The results for the base case are shown in Table 4.

The simultaneous and dynamic approach clearly outperforms any of the other methods. The dynamic approach significantly alters the values of the individual lot sizes (and other production control parame-

Table 4: Optimal lot sizes for the base case with different optimization modes

Lot Size	Optimization Mode	p	$t=1$	$t=2$	$t=3$	π	Evaluations
Static	Iterative	1	16	16	16	47,368	22,896,190
		2	23	23	23	+17.40%	+483%
Static	Simultaneous	1	17	17	17	47,626	4,564,494
		2	26	26	26	+0.54%	-80%
Dynamic	Simultaneous	1	15	67	59	50,143	7,162,194
		2	23	27	35	+5.28%	+57%

ters) in such a way that there is an even larger improvement in overall profit than obtained by the static version. The best static solution in iterative mode (+17.40% when compared to the base case) has been found after 10 cycles, which requires many objective evaluations (+483%). In the next static version, where all the decision variables are simultaneously considered, its profit is slightly improved but it is dramatically more efficient from a computational perspective (-80%). More worthwhile improvements are achieved by the dynamic approach at the expense of more computation time, but still much better than in the iterative mode.

5 Conclusion

An Advanced Resource Planning (ARP) model is developed to support aggregate planning decisions at the intermediate planning horizon, where the uncertainty of the underlying production system is explicitly modeled as a queueing network. This analytical model can be used in a multi-period setting to select the most profitable combination of volume and capacity levels, as well as appropriate lot sizes in complex and stochastic production systems with multiple products, resources and alternative routings. The solution depends on the selling prices and costs for production, inventory, work-in-process, backorders, setups, regular time, overtime and outsourcing.

The ARP application to a simple case problem reveals some interesting insights. DE precisely balances the trade-offs between the different financial terms while taking into account the queueing impact of loading resources that have different process and variability characteristics. Several scenarios for challenges in planning (demand interval variability, finished good cost, labor cost, setup time, setup and process variability, etc.) show that the capacity strategy (regular time, overtime, outsourcing) as well as the assignment strategy of products to routings and periods can be altered in order to compensate for profit reductions or to benefit from profit opportunities.

Scenarios with a positive implication for profit (e.g. lower holding costs) can justify longer product lead times. Other scenarios where capacity is added because it is cheaper (e.g. less expensive labor) can lead to a lead time improvement for one product, but may involve a lead time deterioration for another product because of complex interactions. Setting the lead time offset (delivery time guarantee) is a critical issue: when it is high, all demand is produced in the time bucket where it is due and total available working time can be reduced; when it is low, measures like inventory build-up, extra labor time and outsourcing must be taken to avoid the adverse effects from backorders. Low unit backorder costs create long product lead times, but savings in total labor costs generate more profit. When the outsourcing process is less efficient due to longer delays, overtime is raised to accommodate the larger volume of in-house production and profit is lower.

In general, it is the trade-off between three buffers (inventory, capacity and time) in the ARP model that drives the DE search process towards optimal product-capacity profiles. Each product has its own period dependent lead time with its own specific service level that corresponds to this buffer mix in such a way that profit cannot be improved further. This optimal lead time and delivery performance are embedded into the ARP relationships and can be detected by DE.

An important conclusion is that by modeling the stochastic nature of complex production systems, ARP always significantly outperforms the outcome of a standard aggregate planning model in terms of profit and delivery performance. The structure of the production plan to execute is also completely different (release fractions, lot sizing, labor time, inventory, backorders, outsourcing, etc.). A major difference is that when the ARP planning system is implemented, resources are often not loaded up to the maximum utilization level, but until an appropriate safety capacity is reached. This safety level can vary between resources and planning periods as a function of system dynamics. This is in contrast to traditional planning systems where the safety cushion is fixed and arbitrary because variabilities and uncertainties are not accurately incorporated. Our results proof that the delivery service can be significantly improved when the ARP model is used instead. More profit can be further achieved by optimizing the lot size for each product. Simultaneously considering lot sizes as additional decision variables in the model is preferred over an iterative approach. Dynamically changing them over time has the highest profit potential.

These findings are of great value for practitioners as well: ARP can be used in combination with the DE optimizer to find the best production planning and control parameters in order to guide the S&OP meetings towards profit maximizing decisions with respect to volume, mix, lot sizes and capacity supply while improving their lead time and delivery performance.

Acknowledgement

The authors would like to thank the Fund for Scientific Research Flanders (FWO) for their support with the Post Doctoral Research project for Kris Lieckens.

A Input data ARP sample problem

Table 5: Product data

							Outsource			Demand CH_tD_{pt}		
Product	$c^2_{IA_{pt'}}$	Q_p	fgc_p	boc_p	DW_p	sp_p	$EW_{p(R+1)}$	$\frac{VW_{p(R+1)}}{EW^2_{p(R+1)}}$	osc_p	T1	T2	T3
P1	1	9	12	120	200	250	50	1	120	145.2	132	118.8
P2	1	12	6	60	200	125	50	1	120	290.4	264	237.6

Table 6: Operations data

					Setup			Production		
Product	Routing	Operation	resource	hc_{pro}	SU_{pro}	$c^2_{SU_{pro}}$	suc_{pro}	PR_{pro}	$c^2_{PR_{pro}}$	vc_{pro}
P1	R1	O1	M1	4	1	0.5	80	1	0.5	20
P1	R1	O2	M2	4	0.5	1.5	80	0.5	1.5	20
P1	R1	O3	M4	4	0.3	0.5	80	0.3	0.5	20
P1	R2	O1	M1	4	1	0.5	80	1	0.5	20
P1	R2	O2	M3	4	0.5	0.5	80	0.5	0.5	19
P1	R2	O3	M4	4	0.3	0.5	80	0.3	0.5	20
P2	R1	O1	M4	3	0.8	1.5	120	0.4	1.5	30
P2	R1	O2	M5	3	1	0.5	120	0.5	0.5	30
P2	R2	O1	M1	3	0.8	0.5	120	0.4	0.5	29
P2	R2	O2	M5	3	1	0.5	120	0.5	0.5	30

B Outut data ARP sample problem

Table 7: Resource data that apply to each planning period t

resource	M1, M2, M3, M4, M5
CD_t	31
WD_t	22
RT_{mt}	8
s_m	1
PM_{mt}	0
$MTTR_m$	0
$MTTF_m$	999,999,999
E_m	1
$8 \times rtc_m$	150
$8 \times otc_m$	175

Table 8: Base case

t	$\sum_t \pi_t$	REV_t	$WIPC_t$	SUC_t	FGC_t	BOC_t	OSC_t	LC_t	VC_t
1	16,591	74,212	83	9,852	77	532	0	19,800+1,087	26,190
2	29,176	64,388	73	8,628	0	896	0	16,500+2,459	23,248
3	40,347	59,400	66	7,920	0	880	0	16,500+1,597	21,265
t	CH_tD_{1t}	CH_tD_{2t}	$CH_t\lambda_{1t}$	$CH_t\lambda_{2t}$	EW_{1t}	EW_{2t}	SL_{1t}	SL_{2t}	
1	145.2	290.4	151.65	290.4	92.22	84.13	0.99	0.98	
2	132	264	125.55	264	110.2	85.14	0.96	0.98	
3	118.8	237.6	118.8	237.6	111.33	88.42	0.96	0.98	
t	sh_{1t}+OT_{1t}	sh_{2t}+OT_{2t}	sh_{3t}+OT_{3t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
1	2+0	1+0	1+0	1+0	1+49.71	0	1	0.37	0.63
2	1+56.36	1+0	1+0	1+33.81	1+22.22	0	1	1	0
3	1+46.64	1+0	1+0	1+19.3	1+7.07	0	1	1	0

Table 9: Scenario 1 - Variability of the demand interval

$c^2_{IA_{1t}}$	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
0.25	1	16,337	149.53	90.37	83.52	2+0	1+0	1+49.04	0.35	0.65
0.25	2	29,303	127.67	109.42	85.01	1+56.33	1+33.56	1+22.11	1	0
0.25	3	40,573	118.8	110.94	88.37	1+44.27	1+18.43	1+6.86	1	0
1	1	16,591	151.65	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
1	2	29,176	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
1	3	40,347	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
1.75	1	16,695	152.68	93.54	84.59	2+0	1+0	1+50.18	0.39	0.61
1.75	2	29,050	124.52	110.77	85.24	1+57.41	1+34.32	1+22.36	1	0
1.75	3	40,124	118.8	111.73	88.48	1+48.91	1+20.16	1+7.27	1	0

Table 10: Scenario 2 - Labor cost for regular time

$8 \times rtc_5$	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	$CH_t\lambda_{2t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
150	1	16,591	151.65	290.4	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
150	2	29,176	125.55	264	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
150	3	40,347	118.8	237.6	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
75	1	19,610	145.2	359.82	98.58	59.25	2+0	1+0	2+0	0.38	0.62
75	2	32,753	132	194.58	111.05	84.92	1+65.46	1+8.34	1+0	1	0
75	3	45,574	118.8	237.6	111.32	88.44	1+46.66	1+19.29	1+7.05	1	0

Table 11: Scenario 3 - Efficiency production system

E_5	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	$CH_t\lambda_{2t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
1	1	16,591	151.65	290.4	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
1	2	29,176	125.55	264	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
1	3	40,347	118.8	237.6	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
0.8	1	16,734	145.2	383.66	100.98	82.73	2+3.72	1+0	2+0	0.38	0.62
0.8	2	26,968	132	170.74	110.83	101.11	1+64.89	1+1.34	1+5.4	1	0
0.8	3	37,146	118.8	237.6	111.22	90.57	1+46.47	1+19.76	1+49.61	1	0

Table 12: Scenario 4 - Lead time offsetting

DW_1	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
100	1	29,167	213.47+63.73	68.64	74.22	3+0	1+0	1+43.77	0.35	0.42	0	1
100	2	41,349	118.8	168.16	90.01	1+0	1+15.87	1+23.03	0	1	1	0
100	3	35,705	0	70.11	77.75	1+0	1+0	1+0	0	0	0	1
200	1	16,591	151.65	92.22	84.13	2+0	1+0	1+49.71	0	1	0.37	0.63
200	2	29,176	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	0	1	1	0
200	3	40,347	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	0	1	1	0
300	1	16,821	145.2	132.91	84.55	1+48.98	1+41.93	1+38.67	0	1	1	0
300	2	31,337	132	135.68	87.71	1+32.96	1+25.74	1+23.21	0	1	1	0
300	3	43,550	118.8	138.96	91.3	1+16.83	1+9.58	1+7.77	0	1	1	0

Table 13: Scenario 5 - Inventory cost finished goods

fgc_1	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
6	1	18,019	161.85	98.05	85.96	2+0	1+0	1+50.78	0.44	0.56
6	2	29,248	115.35	112.05	85.47	1+45	1+30.91	1+21.8	1	0
6	3	40,420	118.8	111.32	88.44	1+46.66	1+19.28	1+7.06	1	0
12	1	16,591	151.65	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
12	2	29,176	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
12	3	40,347	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
18	1	15,721	145.2	88.9	83.18	2+0	1+0	1+48.69	0.31	0.69
18	2	29,170	132	109.12	84.94	1+63.35	1+35.58	1+22.44	1	0
18	3	40,341	118.8	111.33	88.43	1+46.64	1+19.29	1+7.07	1	0

Table 14: Scenario 6 - Backorder cost

boc_1	t	$\sum_t \pi_t$	BOC_t	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
60	1	16,179	933	145.2	119.25	83.06	1+62.44	1+46.82	1+38.17	1	0
60	2	30,072	929	132	121.81	86.21	1+45.85	1+30.5	1+22.75	1	0
60	3	41,678	924	118.8	124.79	89.79	1+29.22	1+14.21	1+7.34	1	0
120	1	16,591	532	151.65	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
120	2	29,176	896	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
120	3	40,347	880	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
180	1	15,667	512	145.2	88.9	83.1	2+0	1+0	1+48.81	0.33	0.67
180	2	28,860	880	132	103.67	84.13	1+72.82	1+38.78	1+22.27	1	0
180	3	39,783	861	118.8	105.61	87.6	1+55.99	1+22.45	1+6.9	1	0

Table 15: Scenario 7 - Setup time

SU_{122}	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
0.5	1	16,591	151.65	92.22	84.13	2+0	1+0	1+49.71	0	1	0.37	0.63
0.5	2	29,176	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	0	1	1	0
0.5	3	40,347	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	0	1	1	0
1.25	1	15,808	146.35	92.29	83.59	2+0	1+0	1+49.13	0.18	0.82	0.31	0.69
1.25	2	28,975	130.85	113.35	84.53	1+64.66	1+35.82	1+22.09	0	1	1	0
1.25	3	40,038	118.8	114.97	88.03	1+49.02	1+19.84	1+6.8	0	1	1	0

Table 16: HighRoute1 alternative for Scenario 7 - Comparison of resource characteristics in period T1

	Scenario 7	HighRoute1					
Resource	ρ_{m1}	ρ_{m1}	EWQ_{m1}	VWQ_{m1}	$\tilde{c}'^2_{IA_{m1}}$	$\tilde{c}^2_{IA_{m1}}$	$\tilde{c}^2_{TPRe_{m1}}$
M1	72.76%	↓	↓	↓	↑	↑	↓
M2	8.32%	=	↑	↑	—	↑	=
M3	43.56%	=	↑	↑	—	↑	=
M4	51.59%	↑	↑	↑	↓	↓	↑
M5	75.25%	=	↑	↑	—	↑	=

Table 17: HighRoute1 alternative for Scenario 7 - Comparison of product/routing characteristics in period T1

Product	Route	EW_{pr1}	VW_{pr1}	SL_{p1}
P1	R1	↓	↓	↑
	R2	↑	↓	
P2	R1	↑	↑	↑
	R2	↓	↓	

Table 18: Scenario 8 - WIP cost

hc_{122}	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{21t}	γ_{22t}
0	1	16,615	151.78	92.29	84.15	2+0	1+0	1+49.74	0.38	0.62
0	2	29,186	125.42	110.22	85.14	1+56.21	1+33.78	1+22.21	1	0
0	3	40,360	118.8	111.31	88.44	1+46.67	1+19.29	1+7.06	1	0
4	1	16,591	151.65	92.22	84.13	2+0	1+0	1+49.71	0.37	0.63
4	2	29,176	125.55	110.2	85.14	1+56.36	1+33.81	1+22.22	1	0
4	3	40,347	118.8	111.33	88.42	1+46.64	1+19.3	1+7.07	1	0
8	1	16,565	151.5	92.14	84.11	2+0	1+0	1+49.69	0.37	0.63
8	2	29,166	125.7	110.18	85.13	1+56.51	1+33.86	1+22.22	1	0
8	3	40,334	118.8	111.34	88.43	1+46.61	1+19.29	1+7.06	1	0

Table 19: Scenario 9 - SCV setup time

$c^2_{SU_{112}}$	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
0	1	16,204	149.97	88.84	84.43	2+0	1+0	1+49.96	0.34	0.66	0.33	0.67
0	2	28,872	127.23	109.91	85.08	1+58.18	1+34.30	1+22.27	0	1	1	0
0	3	39,912	118.80	111.33	88.43	1+46.64	1+19.28	1+7.07	0	1	1	0
0.5	1	16,212	150.03	88.94	84.43	2+0	1+0	1+49.95	0.32	0.68	0.33	0.67
0.5	2	28,872	127.17	109.92	85.09	1+58.13	1+34.27	1+22.26	0	1	1	0
0.5	3	39,911	118.80	111.34	88.41	1+46.63	1+19.29	1+7.09	0	1	1	0
1	1	16,229	150.16	89.09	84.42	2+0	1+0	1+49.97	0.3	0.7	0.33	0.67
1	2	28,872	127.04	109.95	85.09	1+57.98	1+34.2	1+22.28	0	1	1	0
1	3	39,911	118.80	111.31	88.44	1+46.67	1+19.30	1+7.05	0	1	1	0

Table 20: Scenario 10 - SCV process time

$c^2_{PR_{112}}$	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{1t}+OT_{1t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
0	1	16,561	152.45	92.28	84.26	2+0	1+0	1+49.77	1	0	0.38	0.62
0	2	28,932	124.75	110.17	85.17	1+55.40	1+33.41	1+22.16	1	0	1	0
0	3	40,002	118.80	111.18	88.43	1+46.57	1+19.12	1+7.05	1	0	1	0
0.5	1	16,357	151.04	90.35	84.42	2+0	1+0	1+49.93	0.84	0.16	0.35	0.65
0.5	2	28,891	126.16	110.16	85.13	1+57.03	1+34.00	1+22.25	1	0	1	0
0.5	3	39,939	118.80	111.35	88.44	1+46.70	1+19.35	1+7.06	1	0	1	0
1	1	16,158	149.60	88.43	84.46	2+0	1+0	1+49.91	0.5	0.5	0.32	0.68
1	2	28,875	127.60	109.85	85.06	1+58.59	1+34.39	1+22.30	0	1	1	0
1	3	39,914	118.80	111.33	88.43	1+46.64	1+19.28	1+7.08	0	1	1	0

Table 21: Scenario 11 - Outsourcing lead time

$EW_{1(2+1)}$	t	$\sum_t \pi_t$	$CH_t\lambda_{1t}$	EW_{1t}	EW_{2t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
40	1	16,336	8.09+137.11	42.65	79.38	1+16.82	0.02	0.03	1	0
40	2	30,074	34.69+97.31	55.75	86.63	1+12.45	0.11	0.15	1	0
40	3	41,268	56.01+62.79	71.96	89.84	1+2.42	0.17	0.30	1	0
50	1	16,183	8.09+137.11	52.09	79.37	1+16.84	0.03	0.03	1	0
50	2	29,809	36.99+95.01	64.38	87.18	1+13.20	0.12	0.16	1	0
50	3	40,926	58.09+60.71	78.88	90.18	1+2.82	0.17	0.32	1	0
60	1	15,979	15.69+129.51	63.57	82.45	1+20.91	0.05	0.06	1	0
60	2	29,464	41.79+90.21	73.94	88.12	1+15.10	0.13	0.19	0.991	0.009
60	3	40,490	61.00+57.80	86.14	90.61	1+3.40	0.18	0.34	1	0

Table 22: Deterministic solution for the base case

t	$\sum_t \pi_t$	REV_t	$WIPC_t$	SUC_t	FGC_t	BOC_t	OSC_t	LC_t	VC_t
1	3,513	72,600	327	9,155	0	15,685	2,363	16,500+268	24,788
2	6,158	66,000	280	8,800	0	14,172	0	16,500+0	23,603
3	8,655	59,400	216	7,920	0	11,059	0	16,500+0	21,209
t	CH_tD_{1t}	CH_tD_{2t}	$CH_t\lambda_{1t}$	$CH_t\lambda_{2t}$	EW_{1t}	EW_{2t}	SL_{1t}	SL_{2t}	
1	145.2	290.4	125.5+19.7	290.4	287	188	0.44	0.66	
2	132	264	132	264	310	161	0.36	0.75	
3	118.8	237.6	118.8	237.6	277	144	0.43	0.79	
t	sh_{1t}+OT_{1t}	sh_{2t}+OT_{2t}	sh_{3t}+OT_{3t}	sh_{4t}+OT_{4t}	sh_{5t}+OT_{5t}	γ_{11t}	γ_{12t}	γ_{21t}	γ_{22t}
1	**1+0**	1+0	1+0	**1+0**	**1+12.22**	0	0.86	0.86	0.14
2	**1+0**	1+0	1+0	1+0	1+0	0	1	0.90	0.10
3	**1+0**	1+0	1+0	1+0	1+0	0	1	0.76	0.24

References

Albin, S. (1981). *Approximating queues with superposition arrival processes*. Ph. D. thesis, Columbia University, Department for Industrial Engineering and Operations Research.

Asmundsson, J., R. Rardin, C. Turkseven, and R. Uzsoy (2009). Production planning with resources subject to congestion. *Naval Research Logistics 56*(2), 142–157.

Bitran, G. and D. Tirupati (1988). Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference. *Management Science 34*(1), 75–100.

Blackstone, J. H. (Ed.) (2010). *APICS Dictionary* (13th ed.). Chicago: APICS The Association for Operations Management.

Buzacott, J. and J. Shanthikumar (1985). On approximate queueing models of dynamic job shops. *Management Science 31*, 870–887.

Buzacott, J. and J. Shanthikumar (1993). *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, NY: Prentice Hall.

Byrne, M. and M. Hossain (2005). Production planning: An improved hybrid ap- proach. *International Journal of Production Economics 93*.

Ettl, M., G. Feigin, G. Lin, and D. Yao (2000). A supply chain network model with base-stock control and service requirements. *Operations Research 48*, 216–232.

Graves, S. (1986). A tactical planning model for a job shop. *Operations Research 34*, 552–533.

Hackman, S. and R. Leachman (1989). A general framework for modeling production. *Management Science 35*, 478–495.

Hahn, G., C. Kaiser, H. Kuhn, L. Perdu, and N. Vandaele (2012). Enhancing aggregate production planning with an integrated stochastic queuing model. Operations Research Proceedings 2011: Springer-Verlag.

Ho, C. (1989). Evaluating the impact of operating environments on mrp system nervousness. *International Journal of Production Research 27*, 1115–1135.

Hopp, W. and M. Spearman (2000). *Factory Physics*. New York: The McGraw-Hill Companies.

Hung, Y. and M. Hou (2001). A production planning approach based on iterations of linear programming optimization and flow time prediction. *Journal of the Chinese Institute of Industrial Engineers 18*(3), 55–67.

Hung, Y. and R. Leachman (1996). A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations. *EEE Transactions on Semiconductor Manufacturing 9*(2), 257–269.

Irdem, D., N. Kacar, and R. UZSOY (2008). An experimental study of an iterative simulation-optimization algorithm for production planning. Miami, FL. 2008 Winter Simulation Conference.

Karmarkar, U. (1987). Lot sizes, lead times and in-process inventories. *Management Science 33*, 409–423.

Karmarkar, U. (1989). Capacity loading and release planning with work-in-progress (wip) and lead-times. *Journal of Manufacturing and Operations Management 2*, 105–123.

Kekre, S. (1984). Some issues in job shop design. Technical report, Simon Graduate School of Business Administration, Rochester, NY, University of Rochester.

Kim, B. and S. Kim (2001). Extended model for a hybrid production planning approach. *International Journal of Production Economics 73*, 165–173.

Lambrecht, M., P. Ivens, and N. Vandaele (1998). ACLIPS: A capacity and lead time integrated procedure for scheduling. *Management Science 44*(11), 1548–1561.

Lieckens, K. and N. Vandaele (2007). Reverse logistics network design with stochastic lead times. *Computers and Operations Research 34*(2), 395–416.

Lieckens, K. and N. Vandaele (2011, August). Differential evolution to solve the lot size problem. Internal Report 1121, K.U.Leuven - Faculty of Business and Economics (download at https://lirias.kuleuven.be/handle/123456789/314744).

Lieckens, K. and N. Vandaele (2012). Multi-level reverse logistics network design under uncertainty. *International Journal of Production Research 50*(1), 23–40.

Missbauer, H. (2002). Aggregate order release planning for time-varying demand. *International Journal of Production Research 40*, 688–718.

Mula, J., R. Poler, J. Garcia-Sabater, and F. Lario (2006). Models for production planning under uncertainty: A review. *Internation Journal of Production Economics 103*.

Nieuwenhuyse, I. V., L. De Boeck, M. Lambrecht, and N. Vandaele (2011). Advanced resource planning as a decision support module for erp. *Computers in Industry 62*(1), 1–8.

Nieuwenhuyse, I. V. and N. Vandaele (2006). The impact of delivery lot splitting on delivery reliability in a two-stage supply chain. *International Journal of Production Economics 104*(2), 694–708.

Nieuwenhuyse, I. V., N. Vandaele, K. Rajaram, and U. Karmarkar (2007). Buffer sizing in multi-product multi-reactor batch processes: Impact of allocation and campaign sizing policies. *European journal of operational research 179*(2), 424–443.

Orlicky, J. (1975). *Material Requirements Planning: the New Way of Life in Production and Inventory Management*. New York: McGraw-Hill.

Pahl, J., S. Voß, and D. Woodruff (2007). Production planning with load dependent lead times: an update of research. *Annals of Operations Research 153*(1), 297–345.

Riano, G. (2003). Transient behavior of stochastic networks: Application to production planning with load-dependent lead times. Technical report, School of Industrial and Systems Engineering, Atlanta, GA, Georgia Institute of Technology.

Shanthikumar, J. and J. Buzacott (1981). Open queueing network models of dynamic job shops. *International Journal of Production Research 19*(3), 255–266.

Shanthikumar, J. and U. Sumita (1988). Approximations for the time spent in a dynamic job shop with applications to due-date assignment. *International Journal of Production Research 26*, 1329–1352.

Srinivasan, A., M. Carey, and T. Morton (1988). Resource pricing and aggregate scheduling in manufacturing systems. Technical report, Graduate School of Industrial Administration, Pittsburgh, PA, Carnegie- Mellon University.

Storn, R. and K. Price (1997). Differential evolution - a simple end efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization 11*, 341–359.

Thomé, A., L. Scavarda, N. Fernandez, and A. Scavarda (2012). Sales and operations planning: A research synthesis. *International Journal of Production Economics 138*.

Thompson, S. and W. Davis (1990). An integrated approach for modeling uncertainty in aggregate production planning. *IEEE Transactions on Systems, Man and Cybernetics 20*, 1000–1012.

Thompson, S., D. Watanabe, and W. Davis (1993). A comparative study of aggregate production planning strategies under conditions of uncertainty and cyclic product demands. *International Journal of Production Research 31*, 1957–1979.

Upasani, A. and R. Uzsoy (2008). Incorporating manufacturing lead times in joint production-marketing models: A review and some future directions. *Annals of Operations Research 161*.

Vandaele, N. and L. De Boeck (2003). Advanced resource planning. *Robotics and computer-integrated manufacturing 19*(1).

Vandaele, N., L. De Boeck, and D. Callewier (2002). An open queueing network for lead time analysis. *IIE Transactions 34*, 1–9.

Vandaele, N., I. Vannieuwenhuyse, and S. Cupers (2003). Optimal grouping for a nuclear magnetic resonance scanner by means of an open queueing model. *European Journal of Operational Research 151*(1), 181–192.

Vollmann, T., W. Berry, D. Whybark, and F. Jacobs (2005). *Manufacturing Planning and Control for Supply Chain Management*. New York: McGraw-Hill.

Voss, S. and D. Woodruff (2003). *Introduction to Computational Optimization Models for Production Planning in a Supply Chain*. New York: Springer.

Whitt, W. (1993). Approximations for the GI/G/m queue. *Production and Operations Management 2*(2), 114–161.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be